DOCUMENT RESUME

ED 049 592 40 EC 032 211

AUTHOR	Walker, Hill M.; Buckley, Nancy K.
TITLE	Assessment and Treatment of Deviant Behavior in Children - Section Four: Evaluation of Three Experimental and One Control Strategy in Facilitating Generalization and Maintenance of Treatment Effects Following Two Months of Treatment in a Token Economy. Final Report.
INSTITUTION	Oregon Univ., Eugene.
SPONS AGENCY	Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
BUREAU NO	BR-6-1308
PUB DATE	70
GRANT	OEG-4-6-061308-0571
NOTE	43p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Behavior Change, *Emotionally Disturbed, *Exceptional Child Research, *Followup Studies, *Program Effectiveness, Program Evaluation, Regular Class Placement, *Reinforcement, Socially Deviant Behavior

ABSTRACT

The document is section four of a six part report on the assessment and treatment of deviant behavior in children. The effects of three experimental strategy were investigated to facilitate generalization of treatment effects following 2 months in a token economy classroom. Forty four subjects were assigned to one of three experimental strategies, peer group reprograming, equation of external stimulus conditions, and teacher training in behavior management techniques, which were implemented in the regular classroom for a 2 month period. The amount of deviant behavior produced by the subjects was the dependent variable. The average for all 44 subjects was felt to show behavior maintenance effects from treatment during followup. The token economy was considered to have a powerful treatment effect. Other aspects of the project are found in section one (EC 032 208) overview; section two (EC 032 209) assessment; section three (EC 032 210) treatment; section five (EC 032 212) teacher behavior; and section six (EC 032 213) single subject experiments. (CD)

ED049592

OE/BEH BR6-1308
PH 40
EC

Final Report

Section Four: Evaluation of Three Experimental and One Control Strategy in Facilitating Generalization and Maintenance of Treatment Effects Following Two Months of Treatment in a Token Economy

Assessment and Treatment of
Deviant Behavior in Children

U. S. O. E. Contract OEG 4-6-061308-0571

Hill M. Walker - Project Co-Director
Nancy K. Buckley - Project Psychologist

EC 032 216

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

PROGRAMMING GENERALIZATION AND MAINTENANCE OF TREATMENT EFFECTS ACROSS TIME AND ACROSS SETTINGS[1]

Submitted for Review and Publication to:
Journal of Applied Behavior Analysis

Hill M. Walker
Nancy K. Buckley
Dept. of Special Education
College of Education
University of Oregon
Eugene, Oregon 97403

ABSTRACT

Effects of three experimental and one control strategy were investigated in facilitating generalization and maintenance of treatment effects following two months in a token economy classroom. At the conclusion of treatment, subjects were randomly assigned to one of three maintenance strategies or a control group and returned to their regular classrooms. The maintenance strategies were peer group reprogramming, equating stimulus conditions between the experimental and regular classrooms, and teacher training in behavior management techniques. The maintenance strategies were implemented in the regular classroom for a two month period and then terminated.

Results indicated a powerful treatment effect produced by the token economy. The average for all 44 subjects showed behavior maintenance effects from treatment during follow-up. The mean percentages of appropriate behavior for the peer group reprogramming and equating stimulus conditions were significantly greater than the control subjects' mean. The teacher training and control group means were not significantly different.

The issue of whether treatment gains, produced by behavior modification techniques, do in fact generalize and maintain across time and across settings has received increasing attention in the last few years (Baer, Wolf, and Risley, 1968; O'Leary, 1969). The amount of research data on this question is limited. However, the available evidence indicates that effective generalization and maintenance of modified behavior does not naturally occur when treatment procedures are abruptly withdrawn (Walker, Mattson, and Buckley, 1969; Birnbrauer, Wolf, Kidder, and Tague, 1965; Kuypers, Becker, and O'Leary, 1968; Patterson, Shaw, and Ebner, 1969). Unless systematic fading procedures are used (O'Leary, Evans, Becker, and Saudargas, 1969) or efforts are made to reprogram the environment in which maintenance is expected (Walker, Mattson, and Buckley, 1969) the probability is substantially reduced that maintenance of the modified behavior will occur automatically or naturally.

There is also evidence to indicate that behavior is situation specific and very responsive to the setting events, reinforcement probabilities and discriminative stimuli that operate in different settings (Patterson and Cobb, 1970). Thus intra-subject behavioral similarity across different settings is probably a function, in part, of the amount of stimulus similarity that exists between such settings. When intervention in one setting alters these controlling variables and behavior change is produced, the failure of the treatment effects to generalize to settings in which these variables

have not been altered is to be expected. Studies reported in the literature indicate that generalization of treatment effects from treatment to nontreatment settings while treatment is on-going is the exception rather than the rule. (Wahler, 1969; Kuypers, Becker, and O'Leary, 1969; Meichenbaum, Bowers, and Ross, 1969; O'Leary, Becker, Evans and Saudargas, 1969; Walker, Mattson and Buckley, 1969).

To date, no studies have been conducted which document long-term maintenance effects following treatment in a token economy in which behavior modification procedures have been used. Studies that have evaluated short term maintenance effects have indicated a need for systematic research on strategies which facilitate generalization and maintenance across time and across settings. The present study investigated the effects of three experimental and one control strategy in facilitating generalization and maintenance of treatment effects following two months of treatment in a token economy.

Method

Subjects

Subjects for the study included 44 male and female subjects in grades three, four, five, and six who were referred to an experimental classroom by the local school district because of academic and behavioral problems experienced in the regular classroom setting. Local schools referred those children who were so disruptive that regular teachers had difficulty accommodating them in the classroom setting. Selection criteria consisted of average or

above average intellectual ability, inadequate academic performance, and socially deviant behavior occuring within the regular classroom setting. All subjects possessed a number of social behaviors that interfered with the learning process. Teacher defiance, non-attending, hyperactivity, and tantrum behavior were attributed to the group as a whole. Individual behaviors exhibited were physical and verbal abuse of peers, rejection of peer initiations, excessive verbal outbursts and non-compliance with teacher instructions and demands. These behaviors were identified as most irritating to the regular classroom teacher. However, the subjects as a group exhibited many additional behaviors illustrative of inadequate social and academic adjustment. All candidates for the experimental classroom were screened with the Walker Problem Behavior Identification Checklist (Walker, 1970) and baseline observations taken in the regular classroom setting prior to placement in the experimental classroom. All subjects scored average or above average on standardized intelligence tests (WISC; Stanford-Binet; CTMM) but had educational deficits in the basic skill areas of reading and math that ranged from one month to three years, nine months below grade level.

Setting

The classroom facilities for the treatment program were adjoining and affiliated with a public elementary school in the local school district. The primary area for academic activities contained six, double desks (approximately 20" x 45" work surface),

the teacher's desk and shelves, and tables for the display of high interest materials. Adjoining rooms provided sink and table facilities for science and art projects, a carpentry room with a variety of tools and wood, and the necessary observation facilities. Space was also available for individual testing, tutoring, and remedial instruction. A small isolation room, for using time-out procedures, adjoined the main classroom area. The children used the same playground and lunch facilities as the regularly enrolled students in the school.

There were 32 elementary schools in the local district. During the course of the study, children were referred and accepted into the experimental classroom from 26 of the 32 elementary schools. The district had 21,700 full-time students and the teacher-pupil ratio was 1:24. This ratio did not include teacher aides, resource teachers, special education teachers, or counselors.

Design

In the fall of 1968, a two year study was begun to investigate three experimental strategies and one control strategy in facilitating generalization and maintenance of treatment effects following two months of treatment in a token economy. The token economy was administered within an experimental classroom setting that accommodated six children at a time. A complete description of the treatment program in the experimental classroom is contained in Walker, Mattson, and Buckley (1969). Over the two year period, a total of 48 subjects received two months of treatment within the token economy. Four groups of six children each received two months of treatment

during each academic year. Four subjects were not included in the study as their parents moved away from the area before the investigation was concluded. At the conclusion of two months of treatment in the experimental classroom, individual subjects within each of the eight treatment groups were randomly assigned to one of three experimental or one control group for purposes of studying maintenance effects. The random assignments were not made until the last week of treatment in the token economy to preclude subjects' receiving differential treatment by the experimenters or teachers due to their placement in a particular experimental maintenance group or a control group. After the random assignments had been made, the six subjects were returned to their respective, regular classrooms. The maintenance and control strategies were implemented for each subject in his regular classroom following treatment in the token economy. The maintenance and control strategies were kept in effect for a two month period in the regular classroom and then terminated. Thus, over the two year period of the study, a total of 44 subjects received two months of treatment in the experimental classroom and were followed up, in either an experimental or control strategy, for a two month period. During the maintenance period following treatment, the 44 subjects were divided into four groups of 11 subjects each. Three of these groups were used for studying three separate maintenance strategies while the fourth served as a control and provided an index of "unprogrammed" generalization and maintenance following treatment.

Procedures

Experimental Group 1 (Peer Group Reprogramming)

In this maintenance strategy, the experimental subject's peer group was reprogrammed so as to support his attempts at appropriate social and academic behavior and to ignore incompatible behaviors such as nonattending, disrupting the class, initiating to peers during study time, etc. The strategy was designed to maintain the subject's post treatment, appropriate behavior in the regular classroom setting by enlisting the support and cooperation of his peer group in helping him control his behavior (Patterson and Anderson, 1964).

When the subject returned to his class, a contingency was implemented in which he had an opportunity to earn points for appropriate social and academic behavior. When the subject earned a predetermined number of points (100), he exchanged them for a group reinforcement for the entire class. Group reinforcements included field trips, cartoon films, class parties, and trips to special school events such as high school basketball games, etc. During the two month maintenance period, the experimental subject had an opportunity to earn points for himself and his classmates during two, 15 minute periods scheduled twice a week. At two fixed times during each week, a staff member would visit the classroom and operate a signalling unit for monitoring the subject's performance and for recording reinforcements. The unit was designed by the project staff and contained two counters, two

stimulus lights, and a buzzer for signalling the experimental subject. The unit was radio controlled and could be operated up to 300 feet away. When the staff member entered the room, he placed the unit on the child's desk so the class as well as the experimental subject could observe the stimulus lights. He then returned to the corner of the room and began operating the unit. The unit contained a green light matched with a plus counter and a red light matched with a minus counter. When the green light was on, it signalled that the experimental subject's behavior was appropriate and that he was earning points on the plus counter. The subject could earn a maximum of 10 points during each 30-minute session. Reinforcements were delivered on a variable interval schedule. Each reinforcement was audible to the experimental subject as he could hear the counter click as it registered one point. Each reinforcing event was visible to the experimental subject and the entire class as the stimulus light temporarily shorted out, causing a flash, whenever a point was registered on the counter. When the subject produced inappropriate behavior, the experimenter extinguished the green stimulus light and activated the red stimulus light. This signalled that the experimental subject's behavior was inappropriate and that he had ten seconds to modify his behavior or he would begin losing points, on a variable interval schedule, on the red counter. Activation of the red light was accompanied by an audible buzzer which signalled to the subject and his classmates that his behavior was inappropriate.

At the end of the 15-minute period a net total of earned points was obtained by subtracting the number of points on the minus counter from the number of points on the plus counter. The total was then announced to the class. This event was usually accompanied by applause from the experimental subject's classmates and by congratulations and praise for his performance. The total was recorded for each session, by a peer, on a thermometer chart located on the wall of the classroom.

The experimental subject had to achieve the right to earn points for himself and his classmates by producing behavior that was acceptable, on the average, to his teacher. Acceptable behavior included exhibiting study behavior, completing assignments on time, not disrupting the class and so on. It was left up to each experimental subject's teacher to determine what was acceptable and what was not acceptable. If the subject produced a protracted piece of inappropriate behavior between the two sessions, the teacher was instructed to notify the project staff and the next scheduled visit by the experimenter would be cancelled. The teacher announced to the class that the next session was cancelled and specified the behavior of the experimental subject that resulted in the cancellation. At this point, the experimental subject began earning the opportunity for the next scheduled session, after the cancelled one.

Each subject's teacher was given a set of instructions for use in explaining the reprogramming strategy to the experimental subject's classmates. The strategy was explained and discussed

immediately prior to the subject's re-entry to the class on a full-time basis. The project staff explained and discussed the strategy with the experimental subject the week he returned to his regular classroom. The instructions to the classroom teacher are presented below.

1. "You, the teacher, explain to the class that _____has been attending an experimental class for the last two months. While there, he has made impressive academic gains and he has learned to work better and complete his assignments on time. We would like to enlist the cooperation of the class in helping him continue this good work in his regular classroom.

2. If they, as a class, learn to ignore his attention getting, disruptive, and non-study behavior and to support his appropriate behavior (listening to instructions, completing assignments, not disrupting the class) then a visitor will come to the class occasionally and operate a box which will signal when _____ has received a point for working well and controlling his behavior.

3. When _____accumulates 100 points, then he exchanges them for a special activity for himself and the entire class. The class can decide what the special activity will be.

4. Emphasize that the visitor will come only if _____'s behavior is appropriate all the time, not just immediately prior to his scheduled visit."

Experimental Group II (Equating Stimulus Conditions)

Subjects in this experimental group were exposed to a strategy, designed to facilitate maintenance, by establishing as many common stimulus elements between the experimental and regular classroom settings as possible. Research evidence in the area of discrimination learning has demonstrated that stimulus similarity or identical elements among components of learning tasks facilitates the process of stimulus generalization. (Butler, 1963; Guttman and

Kalish, 1956; Guttman and Kalish, 1958; Jenkins and Harrison, 1960; Kalish and Guttman, 1957; Kalish and Guttman, 1959). The authors attempted to program generalization and maintenance of treatment effects in this experimental condition, by making the post-treatment environment as similar as possible to the treatment environment in the experimental classroom setting. Thus, a variation of the treatment model administered within the experimental class setting, was implemented for subjects in this condition in their respective, regular classrooms.

The intervention model in the experimental classroom consisted of three primary, treatment variables. These were social reinforcement, token reinforcement, and aversive control procedures in the form of cost contingency and withdrawal from a reinforcing climate. The maintenance strategy for experimental condition II included systematic social reinforcement, token reinforcement, and the academic materials each subject had used during treatment in the experimental classroom. The aversive control procedures were not included in the maintenance strategy since their effective use requires a great deal of close supervision by experienced staff. Time out and cost contingency procedures are also easily abused and used punitively within the regular classroom setting. Thus three sources of stimulus matching were programmed between the experimental and regular classroom settings. These were academic materials, systematic social reinforcement, and token reinforcement.

As in experimental condition I, the project staff explained the maintenance strategy to the subject immediately prior to his

return to his regular classroom. A set of written instructions were designed for the regular classroom teacher to use in implementing the maintenance strategy. These instructions are presented below.

> "For a two month period, you as _____'s teacher, will be provided with the following materials. These materials were used by _____ in the experimental classroom and should be the instructional bases for language, reading, and spelling during this period.

I. Academic Materials (materials varied for each subject)

1. _____

2. _____

3. _____

II. Reinforcer System - You will be provided with 40 blue sheets to be used in recording points. Place a new sheet on the child's desk each day and save the record forms in a folder. _____ has already selected the items he would like to exchange his points for. Each item requires an earned total of 100 points. It is possible for _____ to earn approximately one item per week. The experimenter, Mr. _____ will visit weekly to answer question, give assistance, and bring earned items. The observer, Mr. _____ will visit one or two times per week to take observations on the child only. He has not been told the treatment strategy being used with the child so he will be unable to answer any questions. He will enter the room quietly, sitting at the back of the room and remain only about 30 to 60 minutes.

III. Reinforcement Program - for appropriate social and academic behavior.

1. You are provided with a blue point record form which is divided into 50 squares: (1) 25 points for good academic production (2) 25 points for good student behavior. A maximum of 24 points total can be earned by _____ in any single day.

2. Using a marking pen (felt) award points according to the following achedules:

1. Academics: Check the child's academic work four (4) times a day for correctness, neatness, completion within allotted time, etc. A good time to evaluate is toward the end of bloc activities such as math, reading, spelling, language and so on. If his work is acceptable, satisfactory, or meets stated criteria, administer a maximum of three points per evaluation. Thus, for academics, a child could earn a maximum of twelve (12) points per day. If his work is satisfactory, but barely so, you may want to administer two points or one point instead of the three. If his work is unsatisfactory, you administer no points for that evaluation.

2. Social Behavior: For appropriate social behavior (Examples: paying attention, listening to and following instructions, completing assignments, good playground behavior etc.) evaluate the child four (4) times during the day. For example at 10:15 a.m., 11:45 a.m., 1:30 p.m., and 3:00 p.m. If the child's behavior has been acceptable during the preceding time period, reinforce him with a maximum of three points per evaluation. Thus a child could earn a total of twelve (12) points during the day for appropriate social behavior. If his behavior has been satisfactory, but not exemplary, two points or one point could be administered. If his behavior has not been satisfactory between 10:15 and 11:45, then no points would be awarded at the 11:45 evaluation.

3. Social Reinforcement: It is extremely important that you pair each administration of an earned point with some form of overt, social reinforcement such as praise ("good work" "excellent", or "fine job"), approval, interest, attention and/or affection. In this way, natural social reinforcement from you comes to have the same effect upon behavior as points. However, it is important that you do not verbally reprimand the child when points are withheld."

The project staff met with the classroom teacher prior to the experimental subject's return to the regular classroom and discussed the written instructions. During the session, the staff answered questions about the maintenance procedures and

made sure the teacher understood them. The experimental class teacher explained the programmed, academic materials and demonstrated their use to the regular classroom teacher. The level to which the experimental subject had progressed in the materials was indicated to his teacher and she was instructed to have him continue from that point.

The experimenter who managed the maintenance strategies was introduced at this time and a schedule set up so he could visit the classroom once each week during the two month follow-up period. The experimenter's task was to insure that the teacher implemented the maintenance strategy and to provide whatever assistance the teacher required. However, the experimenter did not provide training in behavior modification techniques to the teachers in this condition.

Since the experimental subjects did not spend the entire school day with their homeroom teachers, a rating system was set up to monitor their performance in other settings within the school setting. The child carried a rating slip with him during the day and other teachers, playground attendents, and the lunchroom manager evaluated his behavior in these settings. Points were awarded by the teacher for these periods, on the basis of ratings the child received.

Experimental Condition III (Teacher Training)

In this condition, the project staff provided the experimental subject's regular classroom teacher with training in behavior

modification techniques in an attempt to facilitate generalization and maintenance of treatment effects. The purpose of this maintenance strategy was to program the classroom teacher to reinforce and support the experimental subject's modified behavior.

In this condition, it was necessary to maintain the teacher's as well as the experimental subject's behavior. Each subject's teacher was contacted and asked if she would like to participate in a maintenance study when he returned to his regular classroom. None of the teachers contacted refused. Requirements of the teacher during the maintenance period were explained and discussed.

Each teacher was enrolled in a division of continuing education class on contingency management. Teachers in this condition attended no formal classes. However, they were required to read and master a semi-programmed text on applications of behavior modification techniques in the regular classroom setting (Buckley and Walker, 1970). The teacher met with the maintenance supervisor and discussed applications of the principles contained in the text. The experimenter provided the teacher with direct training in behavior modification techniques and served as a resource consultant in her application of behavioral principles in maintaining the experimental subject's behavior. After a series of initial training sessions the experimenter visited the class on a weekly basis. The experimenters attempted to control classroom sequencing for the subject so that the training teacher had the child for all basic skills areas.

During the last week the subject was in the experimental class, the regular classroom teacher was invited to spend a day in the experimental class setting. During the morning session, the teacher observed the class from behind one way glass and the project staff explained the treatment program to her. During the afternoon, the teacher worked with the experimental subject on academics and received practice in reinforcing him for appropriate behavior.

At the end of the two month maintenance period, the regular classroom teacher received three hours of credit for the class on contingency management and her tuition was paid by the project. She was informed of these reinforcing consequences when first contacted about participating in the study.

Experimental Group IV (Control)

In this condition, experimental subjects were returned to their regular classrooms, after two months of treatment, with no follow-up support or efforts at programming maintenance. The project staff held a conference with the school and provided results of the subject's academic progress in the experimental classroom. When asked about follow-up support, the project staff indicated that the experimental subject would be observed on a weekly basis for a two month period but that the project's involvement in providing treatment ended when the child returned to his regular classroom.[2]

Staff Time Per Maintenance Group

The ratio of staff time expended to amount of behavior change produced has been discussed by Patterson, Cobb, and Ray (1970) as an important variable in the evaluation of intervention procedures. The experimenters recorded the total amount of staff time required to implement the maintenance strategies for subjects in each of the four groups. Experimental group 1 required an average of four hours of staff time for each subject plus a one hour conference with the school personnel involved. Experimental groups 2 and 3 and the control group required respectively seven hours, nine hours, and zero hours of staff time for implementation. Each group also required a one hour conference with the school.

Observation Procedures

A school observation form, developed by Ray, Shaw, and Patterson (1969) was used to measure defined classroom behavior in this study. The observation form provides for a "...method of 'characterizing' school situations for a given child in such a way as to facilitate understanding the determinants and consequences of social behaviors as well as the relationship of those behaviors to the classroom setting. [p. 1].. The 13 response codes on the form are divided into seven inappropriate and six appropriate categories of classroom behavior. Inappropriate categories include noisy, aggressive, not attending, peer initiation, initiation to peer, movement around the room and inappropriate task.

Appropriate behavior categories include individual work, appropriate group behavior, reciting, volunteering, teacher initiation, and initiation to teacher. Each response code is operationally defined in the manual for the observation form. Criteria are established for each response code along with examples of same.

The form also contains codes for the classroom setting, the social consequences of child behavior and the social agent supplying the consequence. During each six-minute observation session, the activities of the classroom setting are coded as group, individual, transition, or recess. The social consequences of child behavior are coded as no response, attention, praise, compliance, disapproval, non-compliance, and physical (+ or -). The social agent supplying the consequence is coded as teacher, peer, or observer.

The observation form is set up as a grid. Each horizontal line in the grid defines a fifteen-second interval. The six-minute grid is further subdivided into two-minute sections for observer convenience in reading the behavior codes. Using an observation clipboard, set for fifteen-second intervals, the observer moves down one grid line each time he receives a signal from the clipboard. During each fifteen-second interval, the observer records both the behavior of the subject and the social consequences of his behavior by placing the appropriate consequence and agent notation(s) in the space beneath the appropriate behavior code. More than one behavior category can be coded during a single fifteen-

second interval. However, once coded, the same category cannot be recoded during the fifteen-second interval.

A graduate student in school psychology was hired as a research assistant on the project and served as an observer in the regular classroom setting. He recorded baseline and follow-up observations in the regular classroom setting on each group of subjects who received treatment in the experimental class setting. The observer was at no time informed of which subjects were assigned to the various maintenance groups. Practicum students (graduate) in special education were trained to take observations on the experimental subjects during treatment. Observations were taken through one way glass from an observation room adjoining the main classroom.

Reliability

The observer in the regular classroom setting who recorded baseline and follow-up observations was given the coding manual for the observation form developed by Ray et. al. The observer memorized the operational definitions for the response codes and familiarized himself with the grid system, social agent, and consequence codes. After some initial practice in taking observations in the experimental class setting, he accompanied the project training observer to regular classrooms for practice in taking observations on subjects who had been selected for referral to the experimental class. The training observer was experienced in taking observations in both the experimental and regular classroom settings.

Inter-rater reliabilities were calculated by a percent agreement method in which number of agreements was divided by the total number of time intervals. Agreements were defined as two observers coding the same consequence and agent events under the appropriate behavior category in a given fifteen-second interval. The observer was required to reach a criterion of five consecutive two-minute observations of .80 or better with the training observer. Two separate, observation sessions were required to reach criterion in the regular classroom setting. In the first session, 21 observation trials of two minutes duration were completed. Reliabilities during these trials ranged from .25 to 1.00 and averaged .62. In the second session, 12 trials were completed before criterion was reached. Reliabilities between the two observers ranged from .75 to 1.00 during these trials and averaged .93.

Observation data were recorded by practicum students in special education during treatment within the experimental class setting. Observations were taken through one way glass from an observation room adjoining the main classroom. An inter-com system allowed observers to audibly monitor activities in the experimental classroom. Over the two year period, a total of 10 graduate students were trained and used as observers. Each student was trained by the training observer and served a minimum of three months as an observer in the experimental classroom. Each student was required to achieve a criterion of 10 consecutive observations

of .90 or better with the training observer.[3] The observers required an average of 21.7 two-minute observation trials to achieve criterion. Inter-rater reliabilities during these trials averaged .84 and ranged from .12 to 1.00.

Results

Treatment Effects

An analysis of variance for a repeated measures design (Winer, 1962) was used to analyze the baseline, treatment, and follow-up data for all 44 experimental subjects. Scores for each subject were the mean percentage of appropriate behavior produced during baseline, treatment, and follow-up. Each data point in baseline was based on an average of 10, six-minute observations taken in the regular classroom. Treatment scores were based on an average of 64, six-minute observations taken in the experimental classroom. Follow-up scores were based on an average of 32 observations taken in the regular classroom following treatment.

Insert Table 1 About Here

The F ratio of 257.07 in Table 1 indicates the intervention procedures in the experimental classroom produced a very powerful treatment effect. The mean percentage of appropriate behavior for all 44 subjects in baseline was 44.59. During treatment, the mean percentage of appropriate behavior for the same subjects increased to 90.29. During follow-up, the mean percentage was 65.27. The Newman-Keuls test for pair-wise differences was used to examine

all possible differences between baseline, treatment and follow-up means for statistical significance.

Insert Table 2 About Here

The data in Table 2 indicate that all possible mean differences were statistically significant at $p < .01$. Thus, both the treatment and follow-up means were significantly different from baseline. In addition, the treatment and follow-up means were also significantly different.

The treatment program also affected variability in behavioral rates across subjects. For example, during baseline the mean percentage of appropriate behavior varied from 20 to 67. During treatment, it varied from 79 to 99. During follow-up, the range increased and the mean percentage of appropriate behavior varied from 45 to 83. The standard deviations for experimental subjects in baseline, treatment, and follow-up were respectively 13.31, 4.69, and 9.71.

During the first week in the experimental classroom, each subject was given the Stanford Diagnostic Arithmetic Test and the Gates-McKillop Reading Diagnostic Test. During the last week in the experimental classroom, after two months of treatment, each subject was given an alternate form of the same tests.

Insert Table 3 About Here

The F ratio in Table 3 indicates that treatment in the experimental classroom substantially altered the subjects' measured

achievement in math. The mean grade equivalent score for all 44 subjects on the pre-test was 3.5. The mean grade equivalent score for the same subjects on the post-test was 4.5. Achievement ranged from grade level 1.5 to 5.2 on the pre-test and from 1.5 to 6.5 on the post-test.

Insert Table 4 About Here

Inspection of Table 4 reveals that the subjects' measured achievement in reading was also substantially changed. While the effect was not as powerful as that produced in math achievement, it was statistically significant beyond .001. The mean grade level score in reading on the pre-test was 3.6. On the post-test it was 4.5. Mean achievement scores ranged from 1.3 to 7.5 on the pre-test and from 1.8 to 8.6 on the post-test.

Maintenance Effects

Prior to evaluating maintenance effects, it was necessary to test the effectiveness of the randomization procedures used to assign subjects to experimental and control groups following treatment. An analysis of variance for a completely randomized design was used to test for differences among subjects in the four maintenance groups in percentage of appropriate behavior produced during baseline and during treatment. Analyses were also conducted on the subjects' performance on the post-tests in math and reading achievement. F tests were run for the four groups on mean percentage of appropriate behavior produced in

baseline and on mean percentage of appropriate behavior produced during treatment. These F ratios, with 3 and 43 df. were 1.32 for baseline and 1.13 for treatment $p > .05$. Analysis of the four groups' performance on the pre-test and post-test in math achievement yielded F ratios respectively of 2.28 and .92 $p > .05$. The same analysés on the pre-test and post-test data in reading achievement yielded F ratios of 1.01 and .73 $p > .05$. Thus the randomization procedures were effective in removing any systematic differences among the four maintenance groups in percentage of appropriate behavior produced and in measured achievement in math and reading.

An analysis of variance for a completely randomized design was used to analyze maintenance effects among the three experimental groups and one control group. These data were based on an average of 32, six-minute observations taken on the experimental subjects in their regular classrooms during the two-month follow-up period.

Insert Table 5 About Here

The F ratio in Table 5 indicates there were significant differences among the maintenance groups in mean percentage of appropriate behavior produced during the two month follow-up period. A post-hoc-analysis of the means was conducted to isolate the statistically significant mean differences.

Insert Table 6 About Here

Results of this analysis in Table 6 indicate the means of experimental groups one and two were significantly different from the mean of the control group. A mean difference of 7.93 or greater was required for significance at $p \leq .05$. The mean difference of 5 between experimental group 3 and the control group was thus not statistically significant.

An analysis was conducted to determine if the variances of the experimental and control groups were significantly different during the maintenance period. An F-maximum Test for Homogeneity of Variances was applied to both the treatment and the maintenance data.

Insert Table 7 About Here

Inspection of Table 7 indicates the variances of the experimental and control groups were homogeneous during treatment in the experimental classroom setting. However the variances of these same groups were significantly different during the follow-up period and were thus heterogenous. The maintenance procedures thus produced significant differences in the variability as well as the amount of appropriate behavior produced by the experimental subjects in follow-up.

Discussion

As noted earlier, application of the treatment model had a powerful effect across subjects in the amount of appropriate behavior produced and in measured math and reading achievement. The difference between the treatment mean of 90.29 and the

follow-up mean of 65.27 indicates the experimental subjects maintained 72 percent of the average amount of appropriate behavior produced during treatment. When analyzed by maintenance groups, the data indicate that subjects in experimental group 1 (peer group reprogramming) maintained 77 percent of the amount of appropriate behavior produced during treatment. The data for subjects in experimental groups 2 (equating stimulus conditions), 3 (teacher training) and control were respectively 75 percent, 69 percent and 65 percent. There was also an inverse relationship between the amount of staff time invested in the three experimental maintenance groups and the amount of behavior maintenance produced. The experimental group (teacher training) in which the greatest amount of staff time (nine hours) and resources were invested was the least effective in producing maintenance. Conversely, the experimental group (peer group reprogramming) in which the least amount of staff time was invested produced the greatest maintenance.

The two maintenance strategies which were most effective also produced the most reliable maintenance effects. Experimental groups 1 and 2 produced 77 and 75 percent maintenance during follow-up. Their respective variances during this same period were 34.50 and 48.30. The variance for experimental group 3 was 109 and it was 236 for the control group. The greater reliability of the maintenance effects of experimental groups 1 and 2 is reflected in the smaller variances of these groups. Experimental control of maintenance variables appeared to be more

effective in experimental groups 1 and 2 since the authors were in more direct control of the maintenance procedures and were thus able to program and monitor them more carefully. In the peer group reprogramming strategy, individual classroom teachers were directly involved only in giving initial instructions to the class and in contacting the experimenters to cancel a scheduled visit. Thus even an uncooperative teacher would have very little effect upon the results of this maintenance strategy. Three teachers in the peer group reprogramming strategy were judged to be uncooperative by the project staff. Each of the 11 peer groups in this strategy appeared to be quite enthusiastic about the program and were cooperative during its implementation. In the equating stimulus conditions strategy, the experimenters provided the teacher with detailed instructions on how to evaluate and rate the child's performance. Fixed interval times were specified as to when the teacher should reinforce the child for both social and academic behavior. All tangible reinforcers, rating sheets and record forms, and academic materials in the basic skills area were provided for the teacher. In addition, the project staff monitored the child's progress by tracking the number of points earned and by discussions with the teacher. Teachers were thus programmed by the project staff and their execution of the maintenance strategy carefully supervised. However, even with this amount of supervision, several teachers did not use the academic materials provided for the experimental subject. One teacher made points so aversive to the experimental subject that he requested she stop

dispensing them. This strategy also produced a reliable maintenance effect; although the variance of 48.30 was larger than that for the peer group reprogramming strategy.

In contrast, the variances for experimental group 3 and the control group reflect a great deal of inter-subject variability in performance during the maintenance period. The teacher training maintenance strategy was much less effective than either the peer group reprogramming strategy or the equating stimulus conditions strategy in the amount of appropriate behavior produced during maintenance and in the reliability of the maintenance effect. The authors attribute this variability to differences among teachers in the motivation, training, and cooperation necessary to implement the maintenance strategy effectively. In addition, all teachers were given identical levels of training in behavior modification techniques and then supervised in their application of these techniques. The project staff did not suggest specific procedures for the teacher to use; but provided the teacher with support and supervision in those techniques she chose to use. Thus some teachers established token economies for the experimental subject while others relied upon time out procedures and teacher dispensed social reinforcers. These two sources of variation undoubtedly contributed to the differential effectiveness of this strategy in maintaining appropriate behavior following treatment. Some teachers were very responsive to the training procedures and subjects in their classrooms maintained high

levels of appropriate behavior. Other teachers were much less enthusiastic about the maintenance program and were less cooperative in implementing specific maintenance procedures. The amount of appropriate behavior produced by experimental subjects in these classrooms tended to be substantially lower.

There are a number of procedures which might have made the teacher training strategy more effective in producing maintenance. For example, providing more intensive, initial training in behavior modification techniques using modeling, feedback, and hand-shaping procedures would have increased the teachers' skill level. Providing the teacher with video-tapes of her performance during acquisition and providing her with frequency counts on the correct use of reinforcement procedures may have substantially improved her performance. Setting up weekly "data" sessions in which the teacher was informed of the experimental subject's amount of appropriate behavior for that week would have provided support for the teacher's behavior. In addition having each teacher agree to a performance contract in which her grade is contingent upon the amount of behavior maintenance produced could improve performance. A criterion could be established where an average of 75 percent maintenance or more is required for an A grade. An average of 65 to 75 percent would be equivalent to a B grade. An average of below 65 percent would result in an incomplete and possible retraining. There is very little data available on just how much training is necessary to equip the classroom teacher with sufficient skill in behavior modification techniques so she can

apply them effectively in the regular classroom setting. The question of whether the teacher's effective use of these techniques maintains over time is an empirical one. Acquainting the teacher with a new management or teaching technique must be accompanied with the necessary support and supervision when the technique is introduced into the classroom regimen. The teacher is often punished with no change in behavior or changes in the wrong direction when she experiments with behavior modification techniques in the classroom. Unless provided with experienced back up support, the teacher may reject these techniques before she becomes proficient with them.

Subjects who were assigned to the control group showed a great deal of variability in the amount of appropriate behavior produced during follow-up. The variance for the control group during treatment was 21.30. During follow-up, it was 236. The effect of implementing maintenance strategies for the three experimental groups was to establish homogenous conditions across classrooms within each experimental condition. However, control subjects were, in a sense, assigned to 11 different treatments since teachers and school personnel were free to program maintenance in any way they chose. The effect of these heterogenous treatments is undoubtedly reflected in the large variance for the control group.

The results of this study suggest the peer group is a powerful source for programming behavior maintenance following treatment in a token economy. The economy of this technique and the reliability

of its effect establishes its generality as a treatment as well as maintenance procedure. The effect of the equating stimulus conditions strategy is less reliable and it is quite expensive in terms of materials and invested teacher time. The teacher training strategy has a great deal of appeal since it has an obvious multiplier effect. A well trained teacher can apply behavior management techniques to the experimental subject as well as to other members of the class. However, additional research is needed on effective techniques for training teachers in behavior management techniques.

References

Baer, D. M., Wolf, M. M., and Risley, T. Some current dimensions of applied behavior analysis Journal of Applied Behavior Analysis 1968, 1, 91-97.

Birnbrauer, J. S., Wolf, M. M., Kidder, J. D., and Tague, C. E. Classroom behavior of retarded pupils with token reinforcement, Journal of Experimental Child Psychology, 1965, 2, 219-235.

Butter, C. M. Stimulus generalization along one and two dimensions in pigeons. Journal of Experimental Psychology, 1963, 65, 339-345.

Guttman, N. and Kalish, H. I. Discriminability and stimulus generalization. Journal of Experimental Psychology, 1956, 51, 79-88.

Guttman, N. and Kalish, H. I. Experiments in discrimination. Scientific American, Jan. 1958, 198, 78-79.

Jenkins, H. M. and Harrison, R. H. Effect of discrimination training on auditory generalization. Journal of Experimental Psychology, 1960, 59, 246-253.

Kalish, H. I. and Guttman, N. Stimulus generalization after equal training on two stimuli. Journal of Experimental Psychology, 1957, 53, 139-144.

Kalish, H. I. and Guttman, N. Stimulus generalization after training on three stimuli: a test of the summation hypotheses. Journal of Experimental Psychology, 1959, 57, 268-272.

Kuypers, D. S., Becker, W. C., and O'Leary, K. D. How to make a token system fail. Exceptional Children, 1968, 35, 101-109.

Meichenbaum, P. H.,Bowers, K. S., and Ross, R. R. A behavioral analysis of teacher expectancy effect. Journal of Personality and Social Psychology, 1969, 13, 306-316.

O'Leary, K. D. Establishing token reinforcement programs in schools: Issues and problems.. Paper presented in a symposium at the American Psychological Association, Washington, D.C., August 31, 1969.

O'Leary, K. D., Becker, W. C., Evans, M. B., and Saudargas, R. A. A token reinforcement program in a public school: A replication and systematic analysis. Journal of Applied Behavior Analysis, 1969, 2, 3-13.

Patterson, G. R. and Cobb, J. Dyadic analysis of aggressive behaviors: An additional step toward a theory of aggression. In Proceedings of the 6th Minnesota Symposium on Child Psychology. Minneapolis; University of Minnesota, 1970. In press.

Patterson, G. R., Cobb, J. Dyadic analysis of aggressive behaviors: An additional step toward a theory of aggression. In Proceedings of the 6th Minnesota Symposium on Child Psychology. Minneapolis; University of Minnesota, 1970. In press.

Patterson, G.R., Cobb, J. A., and Ray, Roberta S. A social engineering technology for retraining aggressive boys. In H. Adams and L. Unikel (eds.) Georgia Symposium in Experimental Clinical Psychology, Volume II, Pergamon Press, 1970.

Patterson, G. R., Shaw, D. A., and Ebner, M. J. Teachers, peers, and parents as agents of change in the classroom. In F. A. M. Benson (Ed.) Modifying Deviant Social Behavior in Various Classroom Settings. Monograph Series #1, Dept. of Special Education, University of Oregon, Eugene, Oregon, February 1969, 1-48.

Ray, R. S., Shaw, D. A., and Patterson, G. R. Observation in the school: Description of a coding form. Oregon Research Technical Report, 1969.

Wahler, R. G. Setting generality: Some specific and general effects of child behavior therapy. Journal of Applied Behavior Analysis, 1969, 2, 239-246.

Walker, H. M. Walker Problem Behavior Identification Checklist, Test and manual. Los Angeles, Calif: Western Psychological Services, January, 1970.

Walker, H. M., Mattson, R. H., and Buckley, N. K. Special class placement as a treatment alternative for deviant behavior in children. In F. A. M. Benson (Ed.) Modifying Deviant Social Behavior in Various Classroom Settings. Monograph Series #1, Dept. of Special Education, University of Oregon, Eugene, Oregon, February 1969, pp 49-80.

Winer, B. J. Statistical Principles in Experimental Design New York: McGraw-Hill Book Company, 1962.

Footnotes

1. This research was supported by U.S.O.E. Grant OEG 4-6-061308 0571 Assessment and Treatment of Deviant Behavior in Children.

The authors wish to express appreciation to Mr. Tom Hochstatter for his generous cooperation and support during the study. As principal of the school in which the experimental class was located, his administrative skills proved vital to the success of the treatment program.

2. If the school staff asked for suggestions for maintaining the child's behavior, they were told they would have to devise their own program.

3. A more stringent criterion for reliability was required in the experimental class setting since the range and variability of the subjects behavior was substantially reduced by the treatment program e.g. the subjects were producing study behavior a large percentage of the time. Thus reliability among observers tended to attenuate on categories of behavior that had very low frequencies such as noisy, aggressive, and movement. In addition, observers in the experimental classroom changed every three months whereas the same observer was used in the regular classroom setting during follow-up. The experimental classroom observers were less experienced and thus required additional training.

Table 1

Summary of the Repeated Measures
Analysis of Variance of Baseline,
Treatment, and Generalization Scores
For All 44 Experimental Subjects

Source	SS	df	MS	F	p
Total	58722	131			
Subjects	4919	43			
Treatments	46094	2	23,047	257.07	<.001
Error	7709	86	89.65		

Table 2

Newman-Keuls Test of Differences
Between Baseline, Treatment,
and Follow-Up Means

Shapes	b_1	b_3	b_2
Ordered Means	1962	2872	3973

Differences between pairs		b_1	b_3	b_2
	b_1		910	2011
				1101
MS error = 89.65	r=		2	3
Q.99 (r, 86):			3.76	4.28
MS error Q.99 (r, 86):			236.12	268.78

	b_1	b_3	b_2
b_1		*	*
b_3			*

*Mean difference significant beyond .01

Table 3

Summary of the Repeated Measures Analysis of Variance of Pre-Treatment and Post-Treatment Achievement Scores in Math

Source	SS	df	MS	F	p
Total	10,824	87			
Subjects	7,156	43			
Treatments	2,384	1	2,384	79.83	$<.001$
Error	1,284	43	29.86		

Table 4

Summary of the Repeated Measures Analysis of Variance of Pre-Treatment and Post-Treatment Achievement Scores in Reading

Source	SS	df	MS	F	p
Total	27,946	87			
Subjects	22,286	43			
Treatments	1,746	1	1,746	19.18	<.001
Error	3,914	43	91		

Table 5

Summary of the Analysis of Variance of Mean Percentage of Appropriate Behavior for Experimental and Control Groups During the Two Month Follow-up Period

Source	SS	df	MS	F	p
Total	4192	43			
Between groups	763	3	254	2.98	$<.05$
Within groups	3429	40	85		

Table 6

Post Hoc Multiple Mean Comparisons of Experimental and Control Groups' Mean Percentage of Appropriate Behavior Produced During Follow-up

	Exp. Gr. 2 $\bar{x}$=68	Exp. Gr. 3 $\bar{x}$=63	Control Gr. x=59
Exp. Gr. 1 $\bar{x}$=70	2	7	11*
Exp. Gr. 2 $\bar{x}$=68		5	9*
Exp. Gr. 3 $\bar{x}$=63			4

*Critical value of 7.93 required for significance at $p < .05$

Table 7

F-Maximum Test for Homogeneity of Variances
During Treatment and During Follow-up

Group	Variance	F-Maximum Ratio
Treatment		
Exp. Gr. 1	23.80	
Exp. Gr. 2	16.90	$23.80 \div 16.90 = 1.54 > .05$
Exp. Gr. 3	26.10	
Control	21.30	
Maintenance		
Exp. Gr. 1	34.50	
Exp. Gr. 2	48.30	$236 \div 34.50 = 6.94 < .05$*
Exp. Gr. 3	109	
Control	236	

*Critical value of 5.67 required for significance at $p < .05$